



Carnelli, P., Sooriyabandara, M., & Wilson, E. (2019). Large-Scale VANET Simulations and Performance Analysis using Real Taxi Trace and City Map Data. In M. Boban, K. Lin, H-M. Tsai, O. Altintas, C-Y. Wang, & T. Sahin (Eds.), *2018 IEEE Vehicular Networking Conference (VNC 2018): 5-7 December 2018 in Taipei, Taiwan* (Vol. 2018-December). [8628352] (IEEE Vehicular Networking Conference (VNC); Vol. 2018). Institute of Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/VNC.2018.8628352>

Peer reviewed version

Link to published version (if available):
[10.1109/VNC.2018.8628352](https://doi.org/10.1109/VNC.2018.8628352)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via IEEE at <https://ieeexplore.ieee.org/document/8628352>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Large-Scale VANET Simulations and Performance Analysis using Real Taxi Trace and City Map Data

Pietro E. Carnelli*

Toshiba Research EU Ltd. & University of Bristol
Bristol, UK
pietro.carnelli@toshiba-trel.com

Mahesh Sooriyabandara

Toshiba Research EU Ltd.
Bristol, UK
mahesh@toshiba-trel.com

R. Eddie Wilson

Dept. of Engineering Mathematics
University of Bristol
Bristol, UK
re.wilson@bristol.ac.uk

Abstract—Wireless vehicular ad-hoc networks comprised solely of city taxis are investigated for their ability to deliver data across an urban environment. Openly available taxi trace datasets for Rome (Italy) and San Francisco (USA) are combined with respective building footprint and road network topology data from OpenStreetMap, to generate a realistic systems level model of a taxi V2V network. Analysis of LOS and NOLOS constraints on wireless transmission range suggests a minimum threshold of 50m is applicable to ensure LOS in over 90% of cases. Variations in taxi location sampling frequency and filtering techniques for the taxi trace datasets are also investigated. Overall vehicular network performance is computed for an all-to-one transmission scenario for both cities with varying taxi fleet size. Results suggest a non-linear relationship between increases in taxi fleet sizes and the reduction of end-to-end delay; doubling taxi fleet size (using a randomised data folding technique) reduces end-to-end delay by a factor of 0.6–0.7. However, doubling the fleet does not increase the fraction of delivered source messages, which saturates at 0.67–0.71 in most simulations. Finally it appears that taxi networks for delivering messages across urban environments are limited more by their routing than by the number of possible V2V exchanges. In a simulated one-to-all continuous V2V broadcast scenario, over 90% of the taxis within the fleet receive the source message within one hour of the original taxi passing the source node.

Index Terms—VANET, V2V, V2X, network simulator, delay tolerant, sensor networks

I. INTRODUCTION

Vehicular Ad-hoc NETWORKS (VANETs) are an extensively researched area over the last decade, defined by their large scale, rapidly evolving topology and complexity in determining mobility patterns [1]–[3]. The city road network governs the movements of vehicles as well as reducing the potential for radio transmission between nearby vehicles due to obstruction by large roadside buildings.

As such, understanding overall network performance and potential applications of VANETs is an area of current research interest. Current standards such as the IEEE802.11p and LTE-V2V are being developed and tested in real-world environments, but have yet to be widely adopted by vehicle manufacturers. However, the majority of VANET studies are still based on simulation, and are split between modeling radio transmission between moving vehicles (taking into account ray tracing, shadowing, multipath and interference measurements

etc.) and larger scale system-level simulations where hundreds of vehicles are modeled as nodes moving in a city network.

VANET simulators comprised of vehicular mobility and wireless transmission models often couple a discrete event network simulator (such as NS2/3) with a micro-scale traffic simulator such as DIVERT [4]. This allows researchers to explore how safety critical messages propagate between vehicles at complex traffic crossings or how vehicles interact in platoons along highways. However, timestamped geospatial data from real vehicle fleets has recently become available. By combining vehicle trace data with a wireless simulator, large scale VANETs may be simulated and their subsequent performance analysed [5], without the complication of having to model city-wide vehicular traffic.

Our research aims to understand what effect varying VANET fleet size has on the performance of delivering short messages between stationary nodes randomly placed in a city. The set-up is inspired by delay tolerant low power wide area networks that transmit data from sensors across a city to a control centre [6], [7]. Rather than routing data packets across a stationary city-wide node network, we envisage a scenario where vehicles act as access points to store data packets (uploaded as they drive past source/sensor nodes) and forward it via opportunistic V2V exchanges until they reach their sink destination. Note, data packets are not actively routed within the VANET, rather, they are opportunistically exchanged amongst passing taxis until they drive past a sink. At which point, the sink downloads all the data packets stored in the passing taxi.

To understand the performance of our simulated VANETs, we model an all-to-one communication scenario, similar to vehicle content delivery networks research [8]. We simulate an infection-like spread of messages between vehicles (V2V), whilst also imposing a V2I target and multiple source locations where messages are generated. We thus analyse the VANET performance in terms of distributing messages between vehicles and in terms of the network coverage area.

For our simulations, we use real and openly available taxi trace data collected over the course of a month for two cities: Rome (Italy) and San Francisco (USA). Furthermore, we investigate how V2V communication range and LOS versus NOLOS constraints affect performance. We also highlight the effects of different sampling and interpolation rates and the

This research was supported by the EPSRC funded Industrial Doctorate Centre in Systems (Grant EP/G037353/1) and Toshiba Research EU. Ltd.

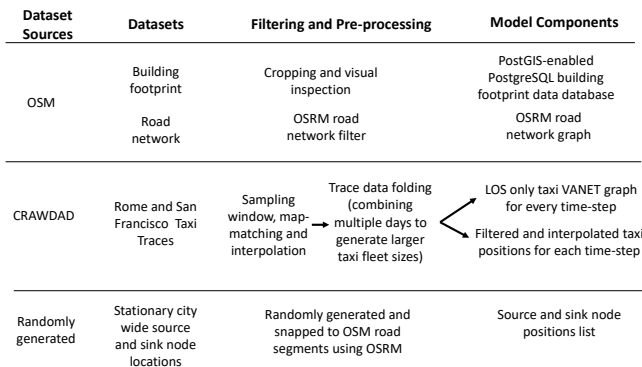


Fig. 1. Overview of our VANET simulator architecture.

importance of filtering the taxi trace data to remove stationary taxis parked at depots. An overview of our modeling system architecture is shown in Fig. 1.

II. RELATED WORK

Simulators such as NS2/3, VEINS and OMNET++ include models of both network protocols and mobility of wireless nodes. VEINS and OMNET++ have plugins to model road networks explicitly, whereas various attempts have been made to combine NS2/3 with vehicular traffic simulators such as DIVERT or SUMO. In any case, both the wireless system and vehicle mobility are modeled using a range of assumptions concerning wireless protocol design, interference patterns, ray-tracing, vehicle density and car-following driver behaviour.

To reduce reliance on pseudo data, open datasets of vehicle mobility in the form of trace data were used in this study. Two cities, Rome (Italy) [9] and San Francisco (USA) [10], provide openly available vehicle trace data, that has already been used in VANET as well as mobility research — see e.g., [5] for a more in depth analysis of the Rome taxi trace data.

Given the inherent complexity of simulating wireless signal propagation in urban environments, large scale VANETs are often modeled with a simple disc-like radius of guaranteed transmission [5], [11]. Given that the reserved portion of the EM spectrum lies within the 5.850–5.925Ghz range, V2V communication without LOS is possible, but challenging in cities given obstructions due to buildings and radio interference. Should VANETs of the future opt for even higher throughput and lower (V2V) delay wireless systems, such as mmWave frequencies, then the LOS criterion becomes crucial for successful transmission of data packets [12].

Recent experiments at the University of Bristol estimate the maximum range at which inter-vehicle wireless transmission is effective in a city environment to be around 200m [13]. V2V experiments conducted in even harsher environments highlight the importance of having LOS when dust clouds are present [14]. Given this background, we investigated the effect of varying the transmission radius range with both NOLOS and LOS assumptions.

III. VANET SIMULATION METHODOLOGY

A. Data Inputs

Open data and open source code were used in our taxi VANET model. Map data for Rome and San Francisco were downloaded from OpenStreetMap (OSM) [15] via the Geofabrik server ([16], [17]) and were cropped according to a central 8×8 km grid (as depicted in Fig. 8) and stored in a PostGIS-enabled PostgreSQL database. OSM data is updated frequently by the community and provided us with road segment and building footprint data. The footprint data is used in assessing and applying the LOS constraint between communicating vehicles. For each vehicle pair within communication range, an ‘in-polygon’ query is made to the building footprint database to test whether or not the communicating vehicles have LOS.

Rome and San Francisco taxi trace data were downloaded courtesy of crawdad.org ([9], [10]). Both taxi trace datasets contain individual taxi identification, GPS locations, and timestamps. Unfortunately no data were available regarding vehicle velocity and uncertainty/error in these measurements. Other taxi trace datasets [11] for cities such as Shanghai and Shenzhen were considered for use, but are not truly open at the time of writing. Open pseudo trace datasets were not considered as real taxi trace data effectively takes into account other vehicular traffic in the city, since participating taxis can only move as fast as the surrounding vehicular traffic. Use of real taxi trace data thus increases the authenticity of our set-up since we do not require traffic simulations that introduce a plethora of hard to verify assumptions.

B. Data Filtering and Pre-Processing

Road network data obtained from OSM was filtered using OSRM (Open Source Routing Machine, [18]). OSRM takes raw OSM data and processes it to develop a graph of the road network. The graph takes into account road lengths, inclines, turn penalties (angles of intersections) and speed limits. This data forms the backbone road network graph for routing and map-matching of the taxi trace data.

In urban environments, where tall buildings and radio interference are commonplace, the accuracy of GPS readings tends to decrease to around 10–15m [19]. This gives problems when interpolating taxi trace data, as vehicles may appear to be driving on one street before disappearing and re-appearing on a neighbouring street, which may sometimes even be disconnected from the former. To overcome this problem, two filters were developed as follows.

The first filter (Filter A in Tab. I) was based on the observed update frequency for the trace datasets, see Fig. 2. As such, sampling windows of 30 and 120 seconds were chosen for Rome and San Francisco respectively, and given 50% overlap with each other to ensure better smoothness of the interpolated taxi trace data. Given the low update frequency (in particular for the San Francisco data), we required a method for interpolating the trace data, such that an active taxi’s location may be estimated at an arbitrary query time. Filter A establishes

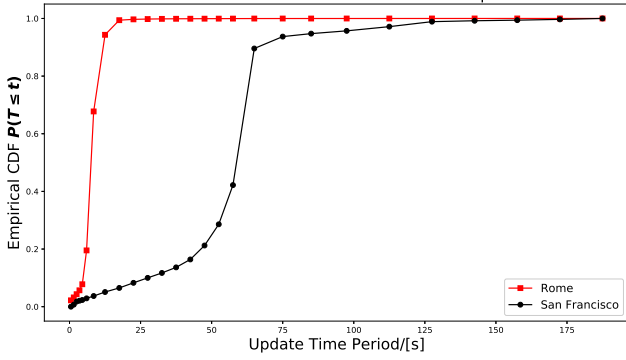


Fig. 2. Empirical cumulative distribution function of raw taxi trace data.

whether a minimum of two data points lie within the query’s sampling window. Should this be the case, the second filter (Filter B in Tab. I) acts on the output of Filter A and map-matches using OSRM’s map-matching system, which is based on a hidden Markov model [20]. Should the two points not lie on the same road segment (but within the sampling window), a ‘most likely’ route (using OSRM’s routing service which neatly takes into account speed limits, turn penalties and traffic direction) is established between the two points before being interpolated. For all taxi trace data interpolations, the start and end times of the interpolated section reflect those given by the raw trace data. Tab. I highlights the percentage of data rejected by each of the two filters and shows the final percentage of the raw trace data that was used (89.5% and 66.1% for Rome and San Francisco respectively). A Python3-OSRM wrapper was used for this research [21].

To prepare data for the VANET simulations, the output of Filters A and B was further interpolated with a discrete time-step of 10s, chosen to balance computational complexity with granularity. The effect of varying the time-step was investigated on a subset of the Rome data, see Fig. 3. Overall, reducing the time-step by a factor of ten (from 10s to 1s) increases the number of message exchange opportunities by a similar factor, suggesting that the time-step is sufficiently small. Fig. 3 also shows how NOLOS and LOS conditions affect the cumulative number of messages exchanged by the VANET over a 24 hour period. For each time-step, two runs were conducted, one with and one without the LOS constraint. The results suggest that the LOS constraint imposes a reduction (on average) of about 32% on the total cumulative messages exchanged in the VANET.

The Euclidean distance between all possible pairs of taxis was evaluated at every 10s for two (San Francisco and Rome) 4 hour taxi trace samples (containing roughly 1000 taxis each). Taxis pairs with a separation distance less than 500m had their LOS condition evaluated by querying our building footprint database. A probability of having LOS given a certain separation distance was evaluated by splitting the results into 50m ‘bins’ before dividing the number of taxi pairs with LOS by the ‘bin’ frequency count. See Fig. 4. The assumption that taxi pairs less than 50m apart have LOS is valid in over 90%

of the cases that we evaluated. However, this rapidly decreases to less than 30% for taxi pair separation distances over 200m, thus highlighting the importance of a sophisticated LOS model to avoid over-estimating VANET message exchanges.

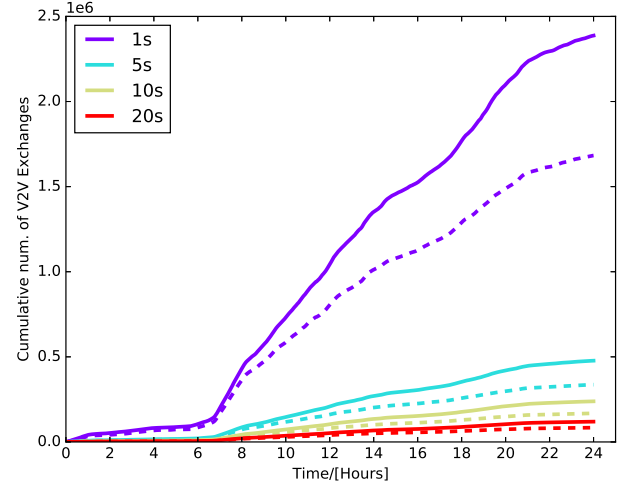


Fig. 3. Cumulative growth in time of V2V message exchange opportunities for various time-steps. Solid lines: all. Dashed lines: LOS only. Computation is for one day of the Rome set-up with an assumed maximum communication range of 200m.

Many taxis spend their time waiting at ranks in airports and train stations [5]. This imposes extra computational costs, since for every queue or group of parked taxis, the VANET model is forced to check for LOS as well as processing V2V message exchanges at each time-step. To avoid unnecessary computational effort and to better assess those active taxis that can form a VANET, two main clusters (one in each city) of stationary taxis were investigated.

In the Rome dataset, taxis queued during the day outside Roma Termini (one of the main train stations). Whereas in San Francisco, taxis were parked at what appears to be a taxi parking lot/depot, located at: $(-122.3947, 37.7516)$. The impact of both stationary clusters were investigated and results for filtering the taxi depot over a randomly selected 12 hour period are shown in Fig. 5. Filtering taxis within 250m from the centre of the San Francisco depot reduced the number of V2V exchanged messages by around a third, far greater than the 10% reduction due the application of the LOS constraint. On the other hand, the impact of filtering stationary taxis outside Roma Termini yielded a reduction of less than 1% of total messages exchanged over the same time period. Consequently taxis outside Roma Termini were not filtered from the dataset.

C. Taxi Trace Data Folding

Variation of the taxi fleet size was achieved by ‘folding’ trace data from randomly selected days, preserving the time of day. To achieve a larger fleet, more days were folded. Due to the different sizes of the participating taxi fleets, different numbers of days had to be folded in order to achieve similar taxi fleet sizes for each city.

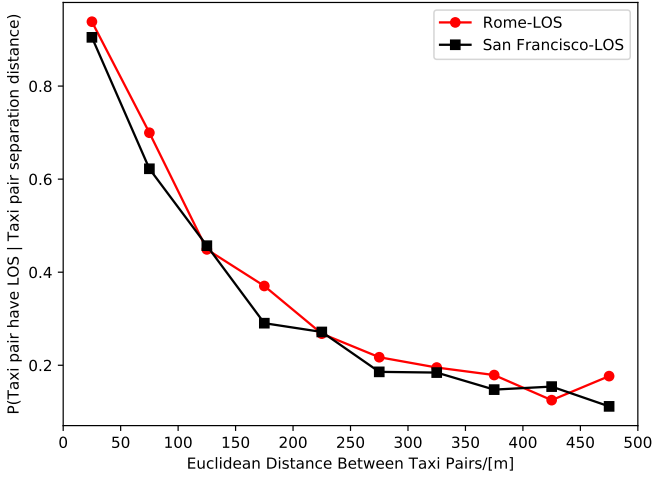


Fig. 4. Likelihood of taxi pairs having LOS given their Euclidean separation distance. Results obtained from sampling four hours of taxi positions at every 10s. Taxi pairs greater than 500m apart were not considered.

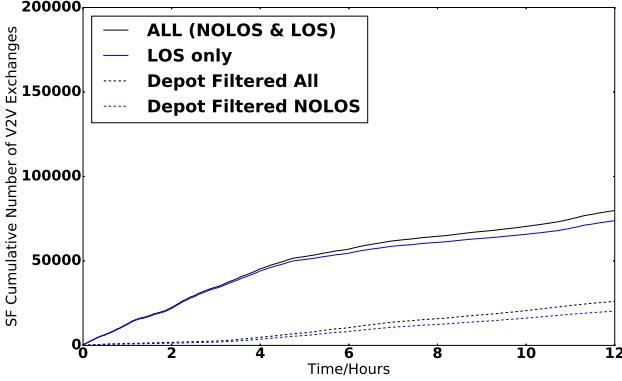


Fig. 5. Effect of filtering the San Francisco taxi depot on cumulative V2V exchanged messages over a randomly selected 12 hour period.

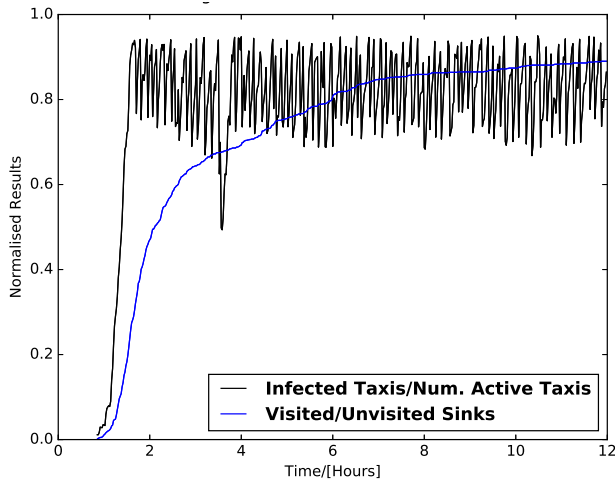


Fig. 6. Proportion of infected taxis and visited sinks during a 12 hour one-to-all VANET simulation for San Francisco. The source is chosen at random from the set of sinks (which are placed on roads), and infects its first taxi at $T \approx 1$ hour.

Once folded, the trace data was pre-processed to establish the new VANETs that emerge as result of combining multiple days of data. This was a computationally intensive task since, for every taxi trace data point, a distance matrix to all other taxis (active at that time-step) had to be computed before an LOS test could be run on all pairs of taxis within the set maximum communication range (200m).

Note, in New York (USA) and London (UK) there are approximately 13,500 [22] and 21,300 [23] licensed taxis, yielding average densities of 13 and 17 taxis/km² respectively. In our simulations, we maintained a similar density by folding 4 and 14 day's worth of taxi traces for San Francisco and Rome receptively yielding a mean of 15 taxis/km².

D. V2X Message Exchange

All our VANET simulations were conducted via a sequence of 10s time-steps. At each time-step:

- An unvisited source-taxi distance matrix was evaluated, in order to assess whether there are taxis within range of an unvisited source. Such taxis are 'infected' with the source's message, which is added to the list of messages that each carries. Visited sources cease to broadcast at future time-steps.
- Taxis within communication range and satisfying LOS (if required) 'infect' each other with the messages that they carry — each taxi's message list becomes the set union of the messages of all taxis involved in the V2V exchange.
- A sink to all active taxis distance matrix is computed. Taxis within range of the sink give their message sets to the sink.
- All new source messages entering the VANET and those received at the sink are timestamped and the number of active taxis forming the VANET is recorded.

E. All-to-One Simulation Runs

Within each 8×8 km study area, 500 nodes were dropped randomly and then snapped to their nearest respective road segments. The pre-computed VANET graphs and filtered and interpolated taxi datasets were then loaded into memory. We then performed an ensemble of 500 all-to-one simulations, where each node plays a turn as the sink, with the other 499 playing as sources. The ensemble procedure prevented results being skewed by outlier source behaviour and allowed us to compute coverage statistics. This whole procedure was repeated for each of the five folded taxi fleets for each city (see Tab. II).

The starting time was randomly selected for each simulation run between 1700 and 1800 hours, as the subsequent 4 hours was deemed the most stable time period in terms of maintaining an average taxi fleet size (see Fig. 9). A random starting time helped support a degree of statistical robustness. Source code for all taxi VANET simulations can be accessed from [24].

IV. TAXI VANET SIMULATION RESULTS

Two measures of delay were used to analyse the performance of the taxi VANETs, total and transit. Total (also

TABLE I
OVERVIEW OF TAXI TRACE DATASETS.

Parameter	Rome, Italy	San Fran- cisco, USA
Data Collection Period	Feb-Mar 2014	May-June 2008
Num. of Participating Taxis	320	536
Median Update Time Period/[S]	7.54	58.7
Sampling Window Size/[S]	30	120
Filter A: % of data rejected due to sampling window	0.92	10.1
Filter B: % of data rejected due to poor location accuracy	9.98	26.5
Overall percentage of data used in simulations (1-A)(1-B)	89.5	66.1

referred to as end-to-end delay) is defined as the time between the simulation starting and when a message (from one of the 499 sources) was delivered to the single sink.

Transit delay measures time elapsed from a source first being visited by a taxi, to when it was delivered to the single sink. Transit delay, by definition, is less than or equal to the total/end-to-end delay.

Fig. 7 shows how the total and transit delay vary with the number of active taxis and those which form a VANET at each time step respectively. In both measures, it is clear that an increase in the number of taxis (as more and more days worth of trace data were folded together) does not result in a proportional decrease in the observed delay. Tab. II further summarises the simulation results and details of the legend used in plotting Figs. 9 & 10. Increasing the mean number of active taxis by a factor of $\simeq 8$ reduces the mean end-to-end delay by nearly a third for both cities.

TABLE II
RESULTS FIGURES LEGEND

Location	Number of Days Folded	Mean Number of Active Taxis	Mean Number of Taxis in VANET	Results Plotting Color, Figs. 9 & 10
San Francisco, USA	1	272	105	Black
	2	523	269	Blue
	4	983	648	Green
	6	1560	1180	Red
	9	2260	1820	Cyan
Rome, Italy	3	258	133	Black
	7	465	277	Blue
	14	983	698	Green
	21	1670	1310	Red
	28	2250	1840	Cyan

The bottom two sub-plots of Fig. 9 show how the number of active taxis and those forming a VANET vary at each time-step. In most cases, the fleet sizes remain fairly constant.

Fig. 9 also highlights the rate at which messages from each of the sources were received over the simulation period. In all simulations, rarely does the number of received messages exceed 65% of the total possible, regardless of the city and the number of days worth of taxi trace data folded. Increasing

the number of taxis seems only to increase the rate at which the number of messages (one per source) were received at the single sink. By and large, within 2 hours of the simulated time period, 60% packet delivery ratio was reached and remained fairly flat from that point.

The large (with respect to fleet/VANET size) standard deviation observed in transit and total delay was further investigated by plotting their respective cumulative distributions for each of the simulation runs in Fig. 10. The plotting legend convention as described in Tab. II is used. Simulations with more than 500 taxis on average achieve a transit and total delay of less than 2,500 seconds (roughly 42 minutes) 80% of the time. Assuming *ceteris paribus*, San Francisco taxi VANETs appear to achieve a slightly better network delay performance than their Rome equivalents.

Fig. 8 shows the locations of the sources/sinks used in the taxi VANET simulations plotted above the road network for both cities. The median transit delay is then plotted in the form of a colour map on top of the source/sink locations. Lighter shades of red indicate lower median delay, in receiving messages from all other (i.e., 499) sinks within the city. As the number of taxis increases, the transit delay decreases: this can be seen both in Fig 7 and in the lighter shades of source/sinks in the bottom two plots of Fig. 8.

There does not appear to be a strong pattern that describes the distribution of transit delay across either city. Sources/sinks which are near or on main roads tend to have lower delays, likely due to the higher probability of a taxi passing by. However, this is a weak trend. The geographical distribution of delay if anything appears to be more correlated with where the taxis were being driven. Preliminary results shown in Fig. 6 suggest that within an hour of a taxi passing a single source, in a simulated one-to-all broadcast scenario (similar to the scenario investigated by [5]), the message is broadcast to nearly 90% of active taxis in the city and this infection level remains fairly constant regardless of taxis dropping in/out of the VANET.

V. DISCUSSION

Overall, increasing the mean size of the taxi fleet decreases the mean delay experienced between transferring messages from one location to another within the bounds of the city. However, the decrease in delay is not inversely linearly proportional to the increase in size of the taxi fleet, as shown in Fig. 7 and briefly summarised in Tab. II.

Network coverage depends strongly on where the taxis drive, since a message can only be relayed to a physical source/sink if the taxi (carrying messages to be delivered) drives within 100m range of the source/sink. Preliminary results in the efficacy of spreading a message within a fleet of taxis (shown in Fig. 6) suggest that the message is received by approximately 90% of the taxi fleet within the first hour of its broadcast. Even with taxis entering and leaving the fleet, the percentage of taxis carrying the message remains fairly constant, varying between 70–90% in an oscillatory fashion. Note the proportion of sinks (out of a total of 1000) visited

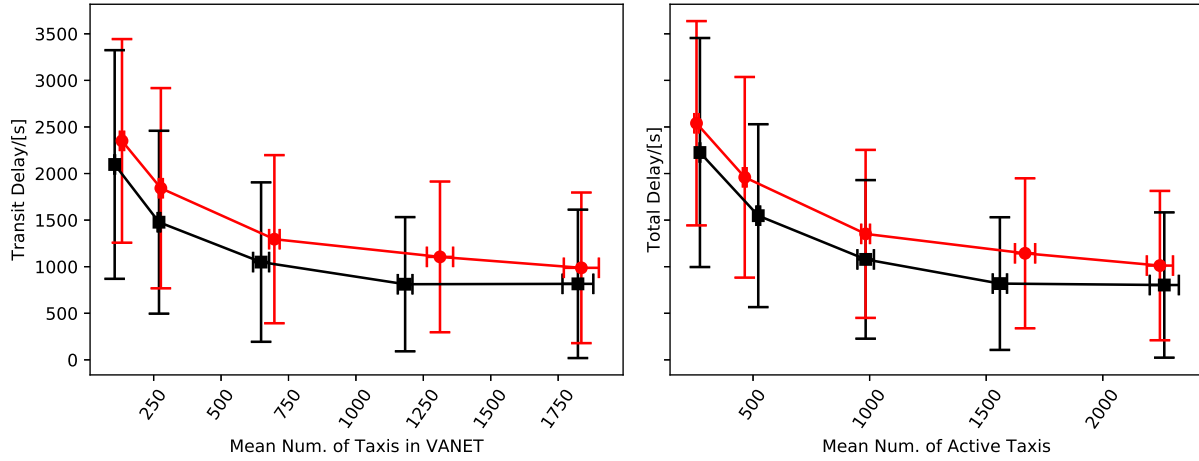


Fig. 7. Mean total and transit delay versus mean number of taxis that are active and those which form a VANET are plotted for different densities of taxis in Rome (red circles) and San Francisco (black squares). Error bars represent the standard deviation from the mean.

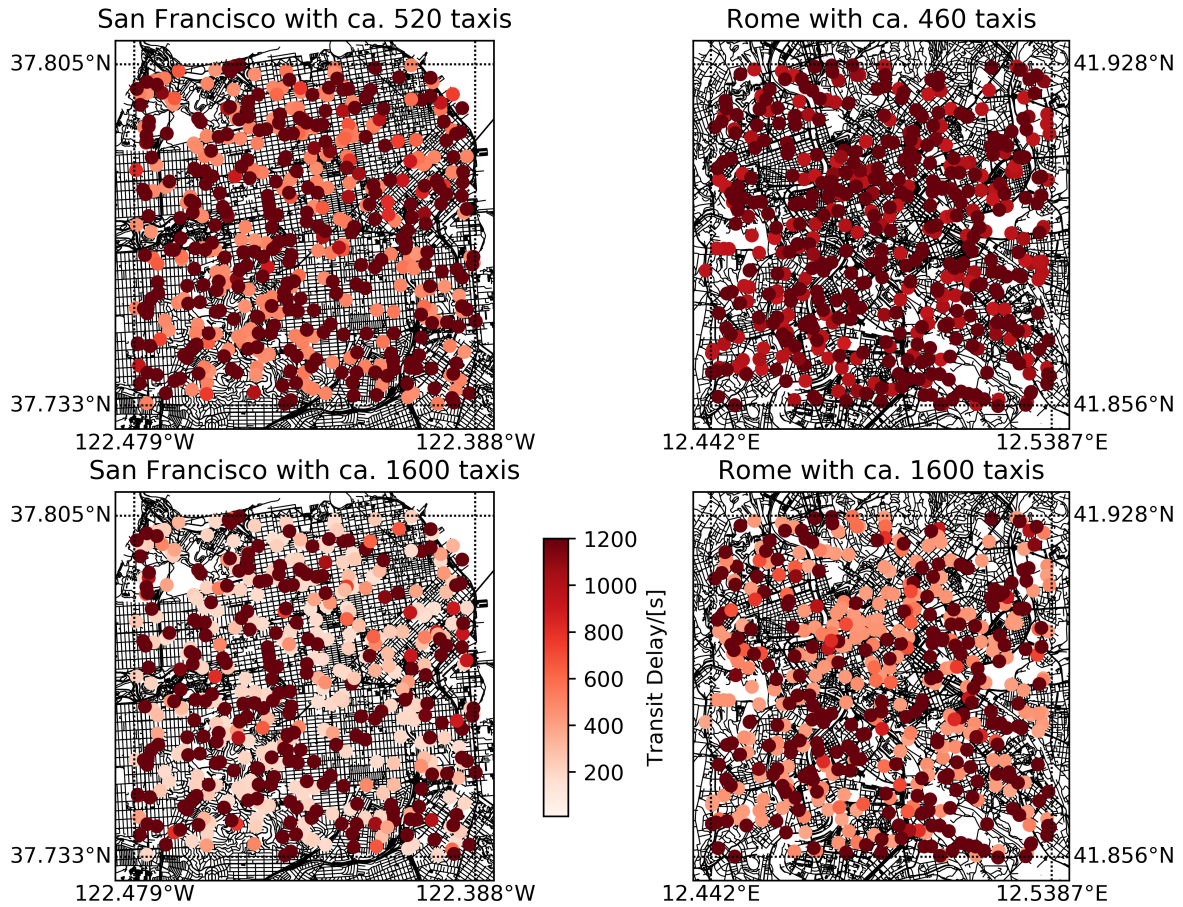


Fig. 8. Transit delay and source locations with road network maps. Latitude and longitude simulation boundaries are highlighted (EPSG:4326 coordinates).

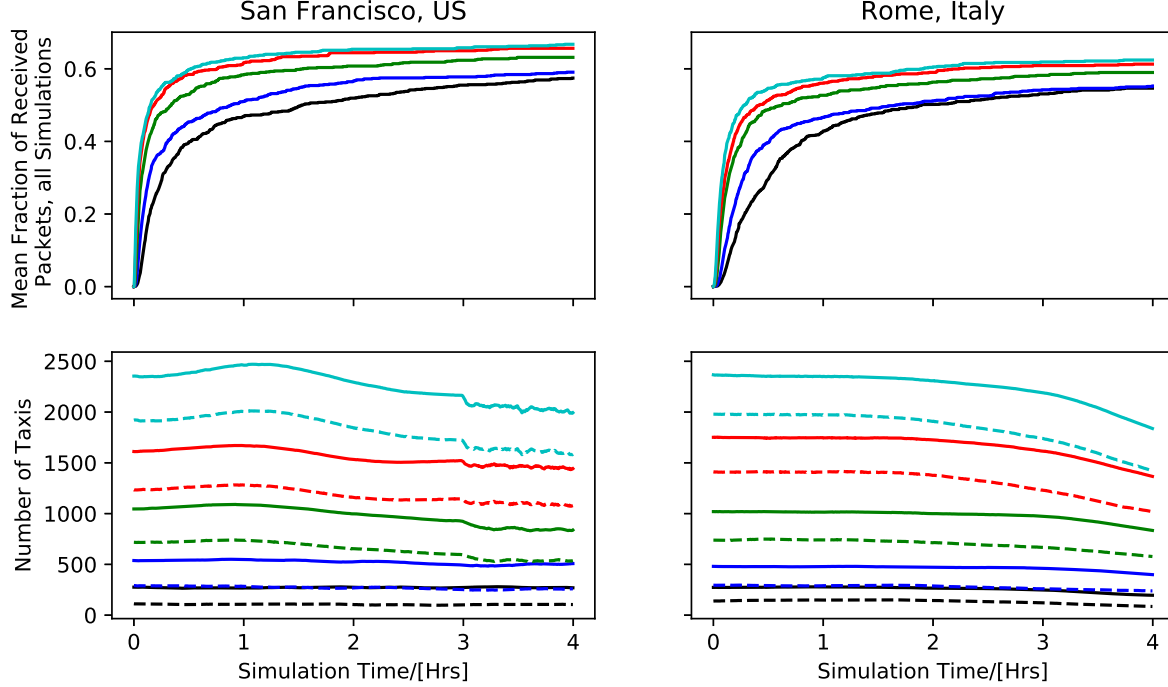


Fig. 9. Fraction of received messages and number of taxis (and those which form a VANET are plotted as dashed lines) are plotted for San Francisco (squares) and Rome (circles) simulations. Colors represent size of folded taxi fleet. See Tab. II for the key.

by the taxi fleet takes longer (5+ hours) to breach the 80% level. This potentially explains why even with greater numbers of taxis in the simulation, it is rare to have a packet delivery ratio greater than 0.7 (as highlighted in Fig. 9). Folding ever more taxi trace data (to increase fleet size) tends to increase the frequency of common trips, rather than the spread/randomness of the origin/destination pairs. Increasing the spread would increase the packet delivery ratio as more sources would likely be visited.

Increasing the length of the simulations might yield higher packet delivery ratios, as over time taxis will cover more ground (highlighted in Fig. 6), hence pass more sources that were not previously visited. However, we quickly hit computational resource limits (principally with memory) running longer simulations or ones with more than 2,500 vehicles.

With regards to our V2V message exchange system, although relatively efficient given the sheer number of V2V exchanges modelled in our simulations, it does have some drawbacks. For example, consider a scenario where one taxi is connected to multiple other taxis (i.e., in a ‘star’ formation where taxis B, C, D and E are connected to a central taxi, A, but not to each other). Depending on the order in which the pairs of V2V exchanges are processed, it is possible to end up with different sets of updated source messages for each of the taxis in question. (An improved algorithm would implement the set union more systematically by iterating within the time-step.) However, as Fig. 6 and Fig. 9 suggest, the messages

spread quite rapidly amongst the active taxi fleet. This suggests that the V2V exchange efficiency is not the limiting factor in reducing end-to-end delay and packet delivery ratios but rather the limiting factor is taxis’ routes (and thus overall ground covered by the fleet during the simulation).

Finally, the use of real trace data has certain drawbacks regarding positional accuracy. Many trace points were rejected due to their seeming implausibility as determined by OSRM’s map-matching algorithm. Thus their location would have been interpreted using OSM’s waypoints and interpolation (ensuring a realistic vehicle profile was set in OSRM) along the road segment. An accuracy of 10–15m may be assumed, effectively the width of most urban roads. Nevertheless, assuming vehicles are roughly positioned in the middle of their respective traffic lane (NB lane direction is taken into account by OSM and OSRM) seems reasonable. On this theme, LOS estimation was limited by the building footprint accuracy, thought to be of the order of a few metres.

VI. CONCLUSION AND FUTURE WORK

We explored two taxi datasets and simulated VANETs in an all-to-one scenario for two cities, Rome and San Francisco. We investigated performance of these networks in terms of end-to-end/transit delay, network coverage, and successful packet delivery. We showed how VANETs could be used to deliver messages across cities as well as showing how their performance was limited primarily by where the taxis were driven/routed, rather than by the number of taxis in the

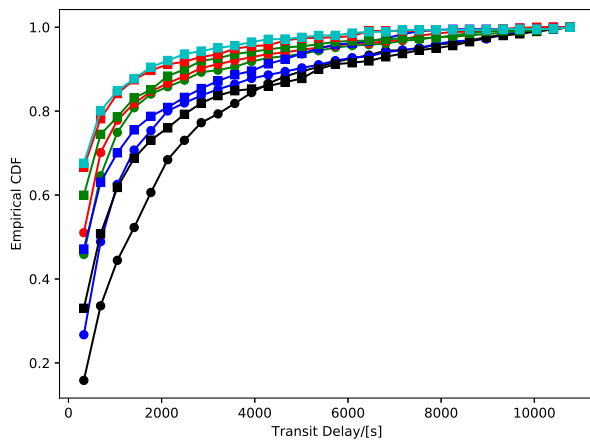


Fig. 10. Empirical cumulative distribution functions of VANET transit delay. Colours represent size of folded taxi fleet. Tab. II describes the plotting legend.

VANET fleet. In terms of future work, ‘chopping’ the taxi trace data, where possible, into trips that can be randomly selected from a pool, could shed light on the geospatial coverage of the VANET. As it stands, even doubling the size of the VANET fleet (in either city) yields small gains both in terms of sources visited (or packets delivered) and reduction of end-to-end delay. Furthermore, we highlighted the importance that a well-defined LOS constraint has on the VANET’s performance. A 100m LOS assumption results in 5% overestimation of V2V communication opportunities, but this reduces to less than 1% with a range of 50m.

To further improve performance, the VANET might include parked vehicles that act as temporary data caches around the city, thus enabling network access to hard-to-reach locations. Other opportunities to improve coverage include better use of empty taxis, which might be re-routed to unvisited sources/sinks.

ACKNOWLEDGEMENT

We thank Huyen Chau Nguyen from Mapbox/OSRM for her patience in aiding us with the OSRM map-matching and routing service.

REFERENCES

- [1] J. Harri, F. Filali, and C. Bonnet, “Mobility models for vehicular ad hoc networks: a survey and taxonomy,” *IEEE Communications Surveys Tutorials*, vol. 11, no. 4, pp. 19–41, Fourth 2009.
- [2] C. Sommer and F. Dressler, “Progressing toward realistic mobility models in vanet simulations,” *IEEE Communications Magazine*, vol. 46, no. 11, pp. 132–137, November 2008.
- [3] H. Wang, R. P. Liu, W. Ni, W. Chen, and I. B. Collings, “Vanet Modeling and Clustering Design Under Practical Traffic, Channel and Mobility Conditions,” *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 870–881, March 2015.
- [4] M. Ferreira, R. Fernandes, H. Conceição, W. Viriyasitavat, and O. K. Tonguz, “Self-organized Traffic Control,” in *Proceedings of the Seventh ACM International Workshop on Vehicular InterNetworking*, ser. VANET ’10. New York, NY, USA: ACM, 2010, pp. 85–90. [Online]. Available: <http://doi.acm.org/10.1145/1860058.1860077>
- [5] R. Amici, M. Bonola, L. Bracciale, A. Rabuffi, P. Loreti, and G. Bianchi, “Performance Assessment of an Epidemic Protocol in VANET Using Real Traces,” *Procedia Computer Science*, vol. 40, pp. 92 – 99, 2014, fourth International Conference on Selected Topics in Mobile and Wireless Networking (MoWNet2014). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877050914014021>
- [6] J. de Carvalho Silva, J. J. P. C. Rodrigues, A. M. Alberti, P. Solic, and A. L. L. Aquino, “Lorawan A low power WAN protocol for Internet of Things: A review and opportunities,” in *2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech)*, July 2017, pp. 1–6.
- [7] U. Raza, P. Kulkarni, and M. Sooriyabandara, “Low Power Wide Area Networks: An Overview,” *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 855–873, Secondquarter 2017.
- [8] J. Eriksson, H. Balakrishnan, and S. Madden, “Cabernet: Vehicular Collivery Using WiFi,” in *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, ser. MobiCom ’08. New York, NY, USA: ACM, 2008, pp. 199–210. [Online]. Available: <http://doi.acm.org/10.1145/1409944.1409968>
- [9] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, “CRAWDAD dataset roma/taxi (v. 2014-07-17),” Downloaded from <https://crawdad.org/roma/taxi/20140717>, Jul. 2014, online: accessed 3-September-2018.
- [10] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, “CRAWDAD San Francisco, USA, taxi trace dataset (v. 2009-02-24),” Downloaded from <https://crawdad.org/epfl/mobility/20090224>, feb 2009, online: accessed 3-September-2018.
- [11] C. Celes, F. A. Silva, A. Boukerche, R. M. d. C. Andrade, and A. A. F. Loureiro, “Improving VANET Simulation with Calibrated Vehicular Mobility Traces,” *IEEE Transactions on Mobile Computing*, vol. 16, no. 12, pp. 3376–3389, Dec 2017.
- [12] A. Tassi, M. Egan, R. J. Piechocki, and A. Nix, “Modeling and Design of Millimeter-Wave Networks for Highway Vehicular Communication,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10676–10691, Dec 2017.
- [13] J. Thota and N. F. Abdullah and A. Doufexi and S. Armour, “Performance of car to car safety broadcast using cellular v2v and ieee 802.11p,” in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, June 2018.
- [14] S. Vernersson, E. Kalpaxidou, and D. Rylander, “Evaluation of wireless short-range communication performance in a quarry environment,” in *2013 International Conference on Connected Vehicles and Expo (IC-CVE)*, Dec 2013, pp. 308–313.
- [15] S. Coast, “Open street map,” 2004, online: accessed 3-September-2018. [Online]. Available: <https://www.openstreetmap.org>
- [16] OSM, “Rome, Italy OSM dataset,” 2007, online: accessed 3-September-2018. [Online]. Available: <http://download.geofabrik.de/europe/italy/centro.html>
- [17] —, “California, USA OSM dataset,” 2007, online: accessed 3-September-2018. [Online]. Available: <http://download.geofabrik.de/north-america/us/california.html>
- [18] D. Luxen, “Project-osrm,” 2010, online: accessed 3-September-2018. [Online]. Available: <https://github.com/Project-OSRM/osrm-backend>
- [19] J. Thota, “MIMO and Raptor codes for cellular V2X,” Ph.D. dissertation, University of Bristol, 2017.
- [20] P. Newson and J. Krumm, “Hidden Markov Map Matching Through Noise and Sparseness,” November 2009, pp. 336–343.
- [21] Ustroetz and PC0179. (2018) Python-OSRM Wrapper, GitHub Repository. Online; accessed 19-October-2018. [Online]. Available: <https://github.com/pc0179/python-osrm>
- [22] NYC Taxi and Limousine Commission, “New York City Taxi and Limousine Commission 2017 Annual Report,” Tech. Rep., March 2017, online; accessed 19-October-2018. [Online]. Available: <http://www.nyc.gov/html/tlc/html/archive/annual.shtml>
- [23] Department for Transport, “Taxi and Private Hire Vehicle Statistics,” Tech. Rep., August 2017, online; accessed 19-October-2018. [Online]. Available: <https://www.gov.uk/government/statistics/taxi-and-private-hire-vehicles-statistics-england-2017>
- [24] Pietro Carnelli. (2018) Taxi VANET simulations, GitHub Repository. Online; accessed 19-October-2018. [Online]. Available: <https://github.com/pc0179/ClusterTaxiVanetSimulations>